

## Small Sample Properties about Estimates in Linear Mixed Models under Linear Restrictions \*

LI ZAIXING

(Department of Mathematics, China University of Mining & Technology, Beijing, 100083)

DING SHENG

(College of Science, China Jiliang University, Hangzhou, 310018)

### Abstract

In the paper, the results for linear models with linear restrictions are partially extended to linear mixed models for longitudinal data with general linear restrictions. At the same time, regularity conditions in Li (2010) were removed and the small sample properties of estimates are investigated.

**Keywords:** Admissibility, longitudinal data, restricted conditions, variance.

**AMS Subject Classification:** 62F10, 62J99.

### §1. Introduction

Li (2010) considered the following linear mixed model (LMM) for longitudinal data under linear restriction,

$$\begin{cases} y_{ij} = x_{ij}^T \beta + z_{ij}^T b_i + \epsilon_{ij}; \\ L\beta = d, \end{cases} \quad (1.1)$$

where  $y_{ij}$  is the  $j^{\text{th}}$  observation for the  $i^{\text{th}}$  subject,  $x_{ij}$  and  $z_{ij}$  are the corresponding covariates, both are observed;  $\beta$  and  $b_i$  are, respectively,  $p$ -dimensional fixed effects and  $q_i$ -dimensional random effects with  $E(b_i) = 0$  and finite covariance matrix  $D_i$ , both unobservable. In the paper, the random error  $\epsilon_{ij}$ , independent of  $b_i$ , are uncorrelated with mean zero and unknown variance  $\sigma^2$ .  $L$  is a  $k \times p$  known matrix and  $d$  is a known  $k \times 1$  vector.

---

\*The project supported by the National Natural Science Foundation of China (11001267) and by the Fundamental Research Funds for the Central Universities in China (2009QS02).

Received August 7, 2010. Revised October 28, 2010.

Denote the subject-specific response vector by  $Y_i = (y_{i1}, \dots, y_{in_i})^\tau$ , and  $X_i, Z_i$  and  $\epsilon_i$  are defined similarly. Model (1.1) can be written as

$$\begin{cases} Y_i = X_i\beta + Z_ib_i + \epsilon_i; \\ L\beta = d. \end{cases} \quad (1.2)$$

Further, denote  $Y = (Y_1^\tau, Y_2^\tau, \dots, Y_m^\tau)^\tau$ , similarly for  $X, b$  and  $\epsilon$ . Also let  $Z = \text{diag}(Z_1, \dots, Z_m)$ . Then, model (1.2) can be written as

$$\begin{cases} Y = X\beta + Zb + \epsilon; \\ L\beta = d, \end{cases} \quad (1.3)$$

and the covariance matrix of  $b$  is  $D = \text{diag}(D_1, \dots, D_m)$ . The linear restrictions in our models is more general than that in Wang (1987) where  $d = 0$ .

The estimation of  $\beta$  is one of the main concerns. Under some regularity conditions, Li (2010) got the estimation and its asymptotic properties. Comparatively, in the paper, we removed these regularity conditions and investigated the existence of the estimates for some functions of  $\beta$ , the corresponding unbiased and variance.

The rest of the paper is as follows: the estimates and their small sample (fixed sample) properties are given in Section 2. The proofs and some detailed derivations are delayed in the Appendix.

Before we close this section, some notations are introduced. For any matrix  $A$ ,  $\|A\|^2$ ,  $\text{tr}(A)$ ,  $\text{Rank}(A)$ ,  $\mu(A)$ ,  $A^-$  and  $A^+$  respectively denote  $A^\tau A$ , the trace of matrix  $A$ , the rank of matrix  $A$ , the space spanned by the column vectors of  $A$ , a generalized inverse matrix  $A$  and the Moore-Penrose generalized inverse matrix. Besides,  $P_A = A(A^\tau A)^-A$  is the projection matrix of  $A$ , and  $P_{A^\perp} = I - P_A$  where  $I$  is the identity matrix and its dimension is the number of rows for  $A$ . For any two nonnegative definite matrices  $A$  and  $B$  which have the same dimension,  $A \leq B$  means that  $B - A$  is a nonnegative definite matrix.

## §2. Estimation for LMMs under Linear Restrictions

According to Li (2010), one considers the following equations

$$\begin{pmatrix} X^\tau P_{z^\perp} X & L^\tau \\ L & 0 \end{pmatrix} \begin{pmatrix} \beta \\ \lambda \end{pmatrix} = \begin{pmatrix} X^\tau P_{z^\perp} Y \\ d \end{pmatrix}. \quad (2.1)$$

By the generalized inverse formula for the bounded matrix (Chapter One in Wang, 1987), one has the solution  $\tilde{\beta}_L$  of  $\beta$  in (2.1). The form of  $\tilde{\beta}_L$  and its property are stated below.

**Theorem 2.1** (i) Equation (2.1) is admissible, and the solution of  $\beta$  is

$$\tilde{\beta}_L = G_{11}X^\tau P_{z^\perp}Y + G_{12}d, \quad (2.2)$$

where

$$\begin{aligned} G_{11} &= T^- - T^-L^\tau Q^-LT^-, & G_{12} &= T^-L^\tau Q^-, \\ T &= X^\tau P_{z^\perp}X + L^\tau L, & Q &= LT^-L^\tau. \end{aligned} \quad (2.3)$$

(ii) The sufficient and necessary condition for  $\tilde{\beta}_L$  to be the least square solution under the restricted conditions is that  $\tilde{\beta}_L$  satisfies equation (2.1).

The proof of Theorem 2.1 is in the Appendix.

**Remark 1** Denote  $\hat{\beta} = (X^\tau P_{z^\perp}X)^+X^\tau P_{z^\perp}Y$ , then

(1) if  $\mu(L^\tau) \subset \mu(X^\tau)$  and  $\text{Rank}(L) = k$ , then

$$\tilde{\beta}_L = \hat{\beta} - (X^\tau P_{z^\perp}X)^+L^\tau(L(X^\tau P_{z^\perp}X)^+L^\tau)^{-1}L\hat{\beta} + T^+L^\tau Q^-d;$$

(2) if  $\text{Rank}(X^\tau P_{z^\perp}) = p$ ,  $\text{Rank}(L) = k$ , then

$$\hat{\beta}_L = \tilde{\beta}_L = \hat{\beta} - (X^\tau P_{z^\perp}X)^{-1}L^\tau(L(X^\tau P_{z^\perp}X)^{-1}L^\tau)^{-1}(L\hat{\beta} - d); \quad (2.4)$$

(3) denote  $\hat{\beta}_{\text{olse}} = (X^\tau X)^{-1}X^\tau Y$  which is an ordinary least square estimator of  $\beta$ , then

$$\text{Cov}(\hat{\beta}_L) \leq \text{Cov}(\hat{\beta}) \leq \text{Cov}(\hat{\beta}_{\text{olse}}); \quad (2.5)$$

the second inequality holds if and only if  $(Z^\tau P_{X^\perp}Z)^{-1} \leq D$ . In other words,  $\hat{\beta}_L$  is the best among these three estimators in the criterion of covariance. (the derivation of the inequality is in Appendix B.)

Similar to the definition in Wang (1987), one can define the concept that “some parameters or their functions are conditionally estimable”.

**Definition 2.1** If there exists some vectors  $a$  and  $a_0$  such that  $E(a^\tau Y + a_0) = c^\tau \beta$  for any  $\beta$  subject to  $L\beta = d$ ,  $c^\tau \beta$  is *conditionally estimable* and then  $a^\tau Y + a_0$  is a conditionally unbiased estimate of  $c^\tau \beta$ .

Note that if  $d = 0$  in model (1.3) and  $a_0 = 0$ , Definition 2.1 is reduced to theirs. Hence, the statement is an extension of theirs. Next, we present a sufficient and necessary condition of conditionally estimable functions in the following theorem.

**Theorem 2.2** Under the restriction  $L\beta = d$ , the sufficient and necessary condition for the fact that  $c^\tau \beta$  is conditionally estimable is that  $c \in \mu(X^\tau : L^\tau)$ .

**Remark 2** The sufficient and necessary condition in Theorem 2.2 is the same as that for linear models with linear restrictions in Wang (1987).

Variances of some estimable functions are investigated in the following theorem.

**Theorem 2.3** (i) If  $c \in \mu(X^\tau P_{z^\perp} : L^\tau)$ , then  $c^\tau \tilde{\beta}_L$  is unique, conditionally linear unbiased estimate of  $c^\tau \beta$  and  $\text{Var}(c^\tau \tilde{\beta}_L) = \sigma^2 c^\tau G_{11} c$  with  $G_{11}$  defined in (2.3);

(ii)  $c^\tau \hat{\beta}_L$  is the unique best linear unbiased estimate of the conditionally estimable function  $c^\tau \beta$  with  $c \in \mu(X^\tau P_{z^\perp} : L^\tau)$ .

Theorem 2.3 (ii) can be partially regarded as an extension of Gauss-Markov theorem in linear models with restrictions (Wang, 1987) to LMMs with restrictions. Harville (1976) considered LMMs without restrictions, and in Harville (1976), the covariance matrix of random effects is assumed to be  $\sigma^2 \Sigma_b$  with known  $\Sigma_b$  and  $\sigma^2$  same to the error variance. Comparatively, covariance matrices of random effects in ours are completely unknown and their connection with error variance  $\sigma^2$  is unknown, either.

Besides, Huang and Lu (2001) considered extended Gauss-Markov theorem for non-parametric mixed-effects models, but the random effects  $b_i$  are not involved in their model. What the so-called random effects in theirs just appear in the representation of the non-parametric function (See Formulas (1.1) and (1.2) on pages 249 and 250 respectively in Huang and Lu, 2001). In other words, if the function is not represented by the mixed models, their problems will not include the random effects part. Furthermore, restrictions are not involved in theirs, either.

Therefore, ours is different from the results in the literature. As far as Theorem 2.3 is concerned, it would be of interest to investigate how the performance of those conditionally estimable function  $c^\tau \beta$  where  $c$  does not belong to the space  $\mu(X^\tau P_{z^\perp} : L^\tau)$ , but  $c \in \mu(X^\tau : L^\tau)$ . A two-step estimation method may be needed and it may be a challenge to get such a best linear unbiased estimates. This deserves further study and it is our ongoing work.

According to Theorem 2.2,  $X\beta$  is conditionally estimable, and then without regularity conditions in Li (2010), the moment-based estimate of  $\sigma^2$  can be defined as follows

$$\tilde{\sigma}_L^2 = \frac{(Y - X\tilde{\beta}_L)^\tau P_{z^\perp} (Y - X\tilde{\beta}_L)}{n - q - t}, \quad (2.6)$$

where  $q = \sum_{i=1}^m q_i$  and  $t = \text{tr}(G_{11} X^\tau P_{z^\perp} X)$  which is equal to  $\text{Rank}(X^\tau P_{z^\perp} : L^\tau) - \text{Rank}(L)$  by Theorem 1.2 on page 4 in Wang (1987). Its property is stated in the theorem below.

**Theorem 2.4**  $\tilde{\sigma}_L^2$  is an unbiased estimate of  $\sigma^2$ .

The proof of Theorem 2.4 is in Appendix A.

### §3. Appendices

#### 3.1 Appendix A: Proofs

Before coming to the proofs, we introduce a needed lemma.

**Lemma 3.1** (1)  $LT^{-}T = L$ ,  $L^{\tau}Q^{-}Q = L^{\tau}$  and  $G_{11}L^{\tau} = 0$ ;

(2)  $G_{11}X^{\tau}P_{z^{\perp}}X + G_{12}L = T^{-}T$ ;

(3)  $P_{z^{\perp}}XG_{11}X^{\tau}P_{z^{\perp}}$  is an idempotent matrix.

(4) In particular, if  $\text{Rank}(X^{\tau}P_{z^{\perp}}) = p$  and  $\text{Rank}(L) = k$ ,  $T$  and  $Q$  are invertible.

**Proof** (1) Since  $\mu(L) \subset \mu(T)$ , we have  $LT^{-}T = L$ . By the derivation for the generalized inverse of the bounded matrix (See Theorem 2.6 on page 12 in Wang, 1987),  $L^{\tau}Q^{-}Q = L^{\tau}$  is obtained. Besides,  $G_{11}L^{\tau} = T^{-}L^{\tau} - T^{-}L^{\tau}Q^{-}Q = 0$ ;

$$\begin{aligned} (2) \quad & G_{11}X^{\tau}P_{z^{\perp}}X + G_{12}L \\ &= T^{-}X^{\tau}P_{z^{\perp}}X - T^{-}L^{\tau}Q^{-}LT^{-}(T - L^{\tau}L) + T^{-}L^{\tau}Q^{-}L \\ &= T^{-}X^{\tau}P_{z^{\perp}}X + T^{-}L^{\tau}Q^{-}QL \\ &= T^{-}T; \end{aligned}$$

(3) By Lemma 3.1 (1) and (2) and noting that  $\mu(X^{\tau}P_{z^{\perp}}) \subset \mu(T)$ , we have

$$(P_{z^{\perp}}XG_{11}X^{\tau}P_{z^{\perp}})^2 = P_{z^{\perp}}X(T^{-}T - G_{12}L)G_{11}X^{\tau}P_{z^{\perp}} = P_{z^{\perp}}XG_{11}X^{\tau}P_{z^{\perp}}.$$

(4) Obvious.  $\square$

**Proof** (Proof of Theorem 2.1) (i) By the admissibility of  $L\beta = d$ , that is  $d \in \mu(LL^{\tau})$ , we have the general solution of  $\beta$ :  $\tilde{\beta}_t = L^{-}d + (I_p - L^{-}L)t$  where  $t$  is an arbitrary  $p \times 1$  vector. Plugging  $\tilde{\beta}_t$  into the first equation in (2.1), we have

$$X^{\tau}P_{z^{\perp}}X(L^{-}d + t - L^{-}Lt) + L^{\tau}\lambda = X^{\tau}P_{z^{\perp}}Y$$

which is admissible. In practice, we only need to verify

$$X^{\tau}P_{z^{\perp}}(Y - XL^{\tau}d) \in \mu(X^{\tau}P_{z^{\perp}}X(I_p - L^{-}L):L^{\tau}) = \mu(X^{\tau}P_{z^{\perp}}X:L^{\tau}).$$

The last equation above that two spaces are the same is derived by the property of linear spaces (see Theorem 1.1 on page 3 in Wang, 1987). This is true because

$$X^{\tau}P_{z^{\perp}}(Y - XL^{\tau}d) \in \mu(X^{\tau}P_{z^{\perp}}X) \subset \mu(X^{\tau}P_{z^{\perp}}X:L^{\tau}).$$

Then the generalized inverse formula for the bounded matrix leads to (2.2).

(ii) Necessity: obvious.

Sufficiency: for arbitrary  $p$  dimensional vector  $a$  which satisfies  $La = d$ , since

$$(P_{z^\perp}X\tilde{\beta}_L - P_{z^\perp}Xa)^\tau(P_{z^\perp}Y - P_{z^\perp}X\tilde{\beta}_L) = (d - d)^\tau P_{z^\perp}(Y - X\tilde{\beta}_L) = 0,$$

we have

$$\begin{aligned}\|P_{z^\perp}(Y - Xa)\|^2 &= \|P_{z^\perp}Y - P_{z^\perp}X\tilde{\beta}_L + P_{z^\perp}X\tilde{\beta}_L - P_{z^\perp}Xa\|^2 \\ &= \|P_{z^\perp}Y - P_{z^\perp}X\tilde{\beta}_L\|^2 + \|P_{z^\perp}X\tilde{\beta}_L - P_{z^\perp}Xa\|^2.\end{aligned}$$

Thus, the proof is complete.  $\square$

**Proof** (Proof of Theorem 2.2) If  $c^\tau\beta$  is conditionally estimable, it is equivalent to the fact: there exists some  $p \times 1$  vector  $a$ , for any  $\beta \in R^p$  subject to  $L\beta = d$ , one has  $a^\tau X\beta + a_0 = c^\tau\beta$ . Since  $L\beta = d$  is admissible, one can select a special solution denoted by  $\beta_0$ . Let  $\beta_* = \beta - \beta_0$ . Then the above statement equals to the following one: for any  $\beta_* \in R^p$ , if  $L\beta_* = 0$ , then  $(a^\tau X - c^\tau)\beta_* = 0$ . In other words,  $(a^\tau X - c^\tau) \in \mu(L^\tau)$ , which is equivalent to the fact that there exist some  $p \times 1$  vector  $e$  such that  $c = X^\tau a + L^\tau e$ .  $\square$

**Proof** (Proof of Theorem 2.3) (i) According to Theorem 2.2,  $c^\tau\tilde{\beta}_L$  is estimable conditionally. The properties of generalized inverse lead to the uniqueness of  $c^\tau\tilde{\beta}_L$ . Since by Lemma 3.1 (2),

$$E(c^\tau\tilde{\beta}_L) = c^\tau(T^-T - G_{12}L)\beta + c^\tau G_{12}d = c^\tau\beta,$$

$c^\tau\tilde{\beta}_L$  is an unbiased estimate of  $c^\tau\beta$ . Besides,

$$\text{Var}(c^\tau\tilde{\beta}_L) = \text{Var}(c^\tau G_{11}X^\tau P_{z^\perp}X\epsilon) = \sigma^2 c^\tau G_{11}X^\tau P_{z^\perp}XG_{11}c = \sigma^2 c^\tau G_{11}c$$

and the last equation is by Lemma 3.1 (1) and Lemma 3.1 (2).

(ii) The unbiased of  $c^\tau\tilde{\beta}$  is verified in Theorem 2.3 (i). Assume that  $a^\tau Y + a_0$  is a conditionally unbiased estimate of  $c^\tau\beta$ . According to Theorem 2.2, there exists some vector  $a_L$  such that  $c = X^\tau P_{z^\perp}a + L^\tau a_L$ . Noting that  $c^\tau G_{11}X^\tau P_{z^\perp}XG_{11}c = c^\tau G_{11}c$  and  $a^\tau P_{z^\perp}XG_{11} = c^\tau G_{11}$  by Lemma 3.1, we have

$$\begin{aligned}&\text{Var}(a^\tau Y + a_0) - \text{Var}(c^\tau\tilde{\beta}_L) \\ &= a^\tau(ZDZ^\tau + \sigma^2 I_n)a - \sigma^2 c^\tau G_{11}c \\ &= a^\tau ZDZ^\tau a + \sigma^2(a - P_{z^\perp}XG_{11}c)^\tau(a - P_{z^\perp}XG_{11}c) \geq 0.\end{aligned}$$

Thus, if and only if  $a = P_{z^\perp} X G_{11} c$ ,  $\text{Var}(a^\tau Y + a_0)$  has the minimal variance  $\text{Var}(c^\tau \tilde{\beta}_L)$ . Then  $\text{E}(a^\tau Y + a_0) = c^\tau \beta$  yields  $a_0 = G_{12} d$ .  $\square$

**Proof** (Proof of Theorem 2.4) Note that

$$\begin{aligned} & (Y - X \tilde{\beta}_L)^\tau P_{z^\perp} (Y - X \tilde{\beta}_L) \\ = & Y^\tau (P_{z^\perp} - P_{z^\perp} X G_{11} X^\tau P_{z^\perp})^2 Y + d^\tau G_{12}^\tau X^\tau P_{z^\perp} X G_{12} d \\ & - 2d^\tau G_{12}^\tau X^\tau P_{z^\perp} (I_n - X G_{11} X^\tau P_{z^\perp}) Y, \end{aligned}$$

and by Lemma 3.1,

$$\begin{aligned} & \text{E}\{d^\tau G_{12}^\tau X^\tau P_{z^\perp} (I_n - X G_{11} X^\tau P_{z^\perp}) Y\} \\ = & d^\tau G_{12}^\tau (T - L^\tau L) \beta - d^\tau G_{12}^\tau X^\tau P_{z^\perp} X (I_n - G_{12} L) \beta \\ = & d^\tau Q^{-1} d - d^\tau d, \\ & d^\tau G_{12}^\tau X^\tau P_{z^\perp} X G_{12} d \\ = & d^\tau Q^{-1} L T^{-1} (T - L^\tau L) T^{-1} L^\tau Q^{-1} d \\ = & d^\tau Q^{-1} d - d^\tau d, \\ & \text{E}\{Y^\tau (P_{z^\perp} - P_{z^\perp} X G_{11} X^\tau P_{z^\perp})^2 Y\} \\ = & \beta^\tau X^\tau P_{z^\perp} X G_{12} L \beta + \sigma^2 \text{tr}(P_{z^\perp} - P_{z^\perp} X G_{11} X^\tau P_{z^\perp}) \\ = & d^\tau Q^{-1} d - d^\tau d + \sigma^2 (n - q - t). \end{aligned}$$

The unbiased of  $\tilde{\sigma}^2$  is obtained.  $\square$

### 3.2 Appendix B

The derivation of (2.5).

Note that

$$\text{Cov}(\hat{\beta}_{\text{olse}}) = \sigma^2 (X^\tau X)^{-1} + \sigma^2 (X^\tau X)^{-1} X^\tau Z (Z^\tau P_{X^\perp} Z)^{-1} Z^\tau X (X^\tau X)^{-1}$$

and

$$\text{Cov}(\hat{\beta}) = \sigma^2 (X^\tau X)^{-1} + \sigma^2 (X^\tau X)^{-1} X^\tau Z (Z^\tau P_{z^\perp} Z)^{-1} Z^\tau X (X^\tau X)^{-1}.$$

Then,  $\text{Cov}(\hat{\beta}) \leq \text{Cov}(\hat{\beta}_{\text{olse}})$  is equivalent to  $(Z^\tau P_{X^\perp} Z)^{-1} \leq D$ . The fact that

$$\text{Cov}(\hat{\beta}_L) = \text{Cov}((I_p - R_{XZL})^\tau \hat{\beta})$$

with  $R_{XZL} = (X^\tau P_{z^\perp} X)^{-1} L^\tau (L (X^\tau P_{z^\perp} X)^{-1} L^\tau)^{-1} L$  yields the first inequality in (2.5).

$\square$

## References

- [1] Harville, D.A., Extension of the Gauss-Markov theorem to include the estimation of random effects, *Ann. Statist.*, **4**(1976), 384–395.
- [2] Huang, S.Y. and Lu, H.S., Extended Gauss-Markov theorem for nonparametric mixed-effects models, *J. Multi. Analysis*, **76**(2001), 249–266.
- [3] Li, Z.X., Estimation in linear mixed models for longitudinal data under linear restricted conditions, *J. Statist. Plann. Inference*, doi:10.1016/j.jspi.2010.08.007.
- [4] Wang, S.G., *Linear Statistical Model and Its Application* (in Chinese), Anhui Education Press, Hefei, China, 1987.

## 带限制线性混合模型中参数估计的小样本性质

李再兴

丁 胜

(中国矿业大学(北京)数学系, 北京, 100083) (中国计量学院理学院, 杭州, 310018)

本文将带有线性限制下的线性模型理论推广至带有一般线性限制下的线性混合效应模型. 同时, 本文在没有李(2010)中的正则条件下, 构造了估计, 考虑了估计的小样本性质.

关键词: 相容的, 纵向数据, 限制条件, 方差.

学科分类号: O212.1.